Comparison of the Anderson-Rubin Test for Overidentification and the Johansen Test for Cointegration

Lennart F. Hoogerheide and Herman K. van Dijk

Econometric and Tinbergen Institutes, Erasmus University Rotterdam, and
Faculty of Economics and Politics, University of Cambridge

Econometric Institute Rapport EI 2001-04

Abstract

In this paper we discuss the similarity between the Anderson-Rubin test for overidentification in a Simultaneous Equations Model and the Johansen test for cointegration in a Vector Autoregressive model. The similar structure of the two models is shown to be important in this respect. An alternative procedure for computing the Anderson-Rubin test is given, which appears to be faster than the conventional method. The derivation of the likelihood ratio test for the hypothesis of reduced rank is given for the general case. Both the Anderson-Rubin test and the Johansen test are shown to be monotonically increasing functions of the singular values of a scaled version of the unrestricted least-squares estimator of the matrix upon which the rank restriction is imposed.
Keywords: Likelihood ratio test, Overidentification, Cointegration, Singular value decomposition.

1 Introduction

In this paper two likelihood ratio tests for the validity of reduced rank restrictions are compared. Reduced rank restrictions occur in several models. Probably, the most well-known models in econometrics are the (incomplete) Simultaneous Equations Model ((IN)SEM) and the Vector Autoregressive (VAR) model, where rank reduction corresponds to the occurrence of overidentification and cointegration, respectively. Two famous tests in this context are the Anderson-Rubin test for overidentification in an (IN)SEM and the Johansen test for cointegration in a VAR model. We discuss the similarities of these tests and the similarities in the mathematical structure of the models. Further, we present the likelihood ratio test for the hypothesis of reduced rank in the general case. Finally, we show that both the Anderson-Rubin test and the Johansen test are monotonically increasing functions of the singular values of a scaled version of the unrestricted least-squares estimator of the parameter matrix upon which the rank restriction is imposed.

The contents of the paper is organized as follows. In Section 2 the similarity of the models to which the tests are applied is discussed. In Section 3 it is shown that the two testing procedures are very similar and an alternative procedure for computing the Anderson-Rubin test statistic

is given, which appears to be faster than the conventional procedure. In Section 4 we present the derivation of the likelihood ratio test for the hypothesis of reduced rank in the general case. In Section 5 it is shown that both the Anderson-Rubin test and the Johansen test can be interpreted as a singular value decomposition. Section 6 concludes.

2 Similarity of the INSEM and the VAR Model

Let the general reduced rank regression model be given by:

$$Z_{0t} = A_1' Z_{1t} + A_2' Z_{2t} + u_t \qquad (t = 1, \ldots, T), \tag{1}$$

where Z_{0t} is an $n_0 \times 1$ vector of endogenous variables, Z_{1t} and Z_{2t} are $n_1 \times 1$ and $n_2 \times 1$ vectors, respectively, of predetermined variables and u_t is an $n_0 \times 1$ vector of disturbances. The difference between Z_{1t} and Z_{2t} is that the model imposes a reduced rank restriction on the $n_1 \times n_0$ matrix A_1, while the $n_2 \times n_0$ matrix A_2 is unrestricted. The u_t's are assumed to be independently and identically $N(0, \Omega)$ distributed disturbances.

Let us consider the (IN)SEM for which only the first structural equation is specified and given by:

$$(y_1 \;\; Y_1) \begin{pmatrix} 1 \\ -\gamma_1 \end{pmatrix} + (X_1 \;\; X_0) \begin{pmatrix} -\beta_1 \\ 0 \end{pmatrix} = \varepsilon_1, \tag{2}$$

where y_1 is a $T \times 1$ vector of observations on the endogenous variable to be explained; Y_1 is a $T \times g_1$ matrix of explanatory endogenous variables; X_1 is a $T \times k_1$ matrix of observations on the predetermined variables that occur in the first structural equation; X_0 (with size $T \times k_0$) contains observations on the exogenous variables that are omitted from the first structural equation. The vectors γ_1 and β_1 contain the structural parameters that we want to estimate. In the case of overidentification the reduced form equations for y_{1t} and Y_{1t} in the (IN)SEM have the shape of the reduced rank regression model (1):

$$\begin{pmatrix} y_{1t} \\ Y_{1t} \end{pmatrix} = \begin{pmatrix} \pi_{10}' \\ \Pi_{10}' \end{pmatrix} X_{0t} + \begin{pmatrix} \pi_{11}' \\ \Pi_{11}' \end{pmatrix} X_{1t} + \begin{pmatrix} v_{1t} \\ V_{1t} \end{pmatrix} \qquad (t = 1, \ldots, T), \tag{3}$$

as the restrictions on the first structural equation of the (IN)SEM impose the reduced rank restriction $\text{rank}(\pi_{10}, \Pi_{10}) \leq g_1$, whereas the matrix (π_{11}, Π_{11}) is unrestricted.

The VAR model can be written in error correction form as

$$\Delta y_t = \Pi' y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i' \Delta y_{t-i} + \Phi' D_t + \varepsilon_t \qquad (t = 1, \ldots, T), \tag{4}$$

where y_t is a $G \times 1$ vector of endogenous variables and D_t is a $D \times 1$ vector of exogenous variables ('deterministic terms'). In the case of cointegration this VAR model in error correction form also has the shape of the model (1), since in this case we must have $\text{rank}(\Pi) \leq r$ for some $r < G$, while $\Gamma_1, \ldots, \Gamma_{k-1}$ and Ψ are unrestricted.

We see the similarity of the (IN)SEM under an overidentification restriction and the VAR model under a cointegration restriction (or VECM).

Both models are instances of the general reduced rank regression model with different variables and parameters playing the same roles, see Table 1.

Model	General model	Reduced form (IN)SEM	VAR in error correction form
Endogenous variables	Z_{0t} $(n_0 \times 1)$	$\begin{pmatrix} y_{1t} \\ Y_{1t} \end{pmatrix}$ $((g_1+1)\times 1)$	Δy_t $(G \times 1)$
Predetermined variables corresponding to parameter matrix with reduced rank	Z_{1t} $(n_1 \times 1)$	X_{0t} $(k_0 \times 1)$	y_{t-1} $(G \times 1)$
Predetermined variables corresponding to unrestricted parameter matrix	Z_{2t} $(n_2 \times 1)$	X_{1t} $(k_1 \times 1)$	$\begin{pmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-k+1} \\ D_t \end{pmatrix}$ $((G(k-1)+d)\times 1)$
Disturbances	u_t $(n_0 \times 1)$	$\begin{pmatrix} v_{1t} \\ V_{1t} \end{pmatrix}$ $((g_1+1)\times 1)$	ε_t $(G \times 1)$
Parameter matrix with reduced rank	A_1 $(n_1 \times n_0)$	$(\pi_{10}\ \ \Pi_{10})$ $(k_0 \times (g_1+1))$	Π $(G \times G)$
Unrestricted parameter matrix	A_2 $(n_2 \times n_0)$	$(\pi_{11}\ \ \Pi_{11})$ $(k_1 \times (g_1+1))$	$\begin{pmatrix} \Gamma_1 \\ \vdots \\ \Gamma_{k-1} \\ \Phi \end{pmatrix}$ $((G(k-1)+d)\times G)$

Table 1: Roles played by the variables and parameter matrices in the general reduced rank regression model, the reduced form of the (IN)SEM under an overidentification restriction and the VAR model under a cointegration restriction.

3 Similarity of the Testing Procedures

The similarity of the reduced form of the (IN)SEM and the VAR model in error correction form suggests that the procedures for obtaining the Anderson-Rubin test statistic and the Johansen test statistic might also be similar. These procedures are given in appendix A. We might expect that we could compute the Johansen test statistic by the procedure that is commonly used for performing the Anderson-Rubin test, i.e. we might expect that an alternative procedure for computing the Johansen test statistic is given by:

1. Regress Δy_t on $\Delta y_{t-1}, \ldots, \Delta y_{t-k+1}$, and D_t, and obtain the residual vectors e_t^* ($t = 1, \ldots, T$). The sample covariance matrix of these residuals is

$$\hat{\Omega}^* \equiv \frac{1}{T} \sum_{t=1}^{T} e_t^* e_t^{*'}.$$

2. Regress Δy_t on $\Delta y_{t-1}, \ldots, \Delta y_{t-k+1}$,$D_t$ and y_{t-1}, and obtain the residual vectors e_t. The sample covariance matrix of these residuals is

$$\hat{\Omega} \equiv \frac{1}{T} \sum_{t=1}^{T} e_t e_t'.$$

3. Determine the eigenvalues l_i ($i = 1, \ldots, G$) of the matrix

$$\hat{\Omega}^{-1/2} \hat{\Omega}^* \hat{\Omega}^{-1/2}, \tag{5}$$

   where $1 \leq l_1 \leq l_2 \leq \ldots \leq l_G$. Then the likelihood ratio test statistic is

$$LR = T \sum_{i=1}^{G-r} \log(l_i). \tag{6}$$

We will show that this alternative procedure indeed yields the same test statistic as the procedure in appendix A on page 10. This result is formalized in the following theorem:

Theorem 1. *The test statistic* (6) *is the same function of the data as the conventional Johansen test statistic* (20) *in Appendix A.*

The proof follows below, but first we give the following lemma that we use in the proof.

Lemma 1. *The matrices AB and BA have the same nonzero eigenvalues.*
If A and B are square matrices (of the same order), the multiplicity of the eigenvalue 0 *is also the same for the matrices AB and BA.*
If A is $p \times q$ and B is $q \times p$ with $p > q$, AB has the same eigenvalues as BA plus $p - q$ more times the eigenvalue 0.
Moreover, if $\lambda \neq 0$ is an eigenvalue of AB corresponding to the eigenvector v, then λ is an eigenvalue of BA corresponding to the eigenvector Bv.

The proof of Lemma 1 is left to the reader; the idea behind the proof is that premultiplying $ABv = \lambda v$ by the matrix B yields the equality $BA(Bv) = \lambda Bv$, where Bv can only be the zero vector if $\lambda = 0$.

Proof of Theorem 1: We can factorize the matrix (5) as

$$\hat{\Omega}^{-1/2}\hat{\Omega}^{*}\hat{\Omega}^{-1/2} = AB$$

with

$$A = \hat{\Omega}^{-1/2}\left(\hat{\Omega}^{*}\right)^{1/2} \quad \text{and} \quad B = \left(\hat{\Omega}^{*}\right)^{1/2}\Omega^{-1/2}.$$

From Lemma 1 we have that the eigenvalues l_i $(i = 1, \ldots, G)$ of (5) are exactly the eigenvalues of the matrix

$$BA = \left(\hat{\Omega}^{*}\right)^{1/2}\hat{\Omega}^{-1}\left(\hat{\Omega}^{*}\right)^{1/2}, \tag{7}$$

since A and B are square. This implies that the eigenvalues of the inverse of (7), the matrix

$$\left(\hat{\Omega}^{*}\right)^{-1/2}\hat{\Omega}\left(\hat{\Omega}^{*}\right)^{-1/2}, \tag{8}$$

are equal to $1/l_i$ $(i = 1, \ldots, G)$.

Now notice that

$$\hat{\Omega}^{*} = S_{00},$$

since the vectors e_t^{*} $(t = 1, \ldots, T)$ on page 4, with sample covariance matrix $\hat{\Omega}^{*}$, have exactly the same definition as the vectors r_{0t} $(t = 1, \ldots, T)$ in appendix A on page 10, with sample covariance matrix S_{00}. Further notice that

$$\hat{\Omega} = S_{00} - S_{01}S_{11}^{-1}S_{10}, \tag{9}$$

since the residuals in the regression of the unrestricted error correction form (4), with sample covariance matrix $\hat{\Omega}$, are the same as the residuals in the *partial regression* of r_{0t} on r_{1t}, with sample covariance matrix $S_{00} - S_{01}S_{11}^{-1}S_{10}$. That these residuals are the same can easily be derived from the Frisch-Waugh theorem for partial regression. Therefore, the matrix (8) is equal to

$$\begin{aligned}\left(\hat{\Omega}^{*}\right)^{-1/2}\hat{\Omega}\left(\hat{\Omega}^{*}\right)^{-1/2} &= S_{00}^{-1/2}(S_{00} - S_{01}S_{11}^{-1}S_{10})S_{00}^{-1/2} \\ &= I - S_{00}^{-1/2}S_{01}S_{11}^{-1}S_{10}S_{00}^{-1/2}.\end{aligned}$$

This implies that the eigenvalues $1/l_i$ of (8) are equal to $1-m_i$, where the m_i's are the eigenvalues of the matrix

$$S_{00}^{-1/2}S_{01}S_{11}^{-1}S_{10}S_{00}^{-1/2}, \tag{10}$$

which are by Lemma 1 exactly the eigenvalues of the matrix (19) in appendix A, since we can factorize the matrix (10) as

$$S_{00}^{-1/2}S_{01}S_{11}^{-1}S_{10}S_{00}^{-1/2} = A^{*}B^{*}$$

with

$$A^{*} = S_{00}^{-1/2}S_{01}S_{11}^{-1/2} \quad \text{and} \quad B^{*} = S_{11}^{-1/2}S_{10}S_{00}^{-1/2}, \tag{11}$$

where A^{*} and B^{*} are square matrices, so that

$$B^{*}A^{*} = S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2} \tag{12}$$

is the matrix (19). We have now shown that the l_i's in the procedure on page 4 and the m_i's in the procedure in appendix A on page 10 are related as follows:

$$\frac{1}{l_i} = 1 - m_i \quad \Leftrightarrow \quad l_i = \frac{1}{1-m_i}, \tag{13}$$

since they are eigenvalues of the same matrices and since the function $1/(1-x)$ is a monotonically increasing function of x for $0 < x < 1$, so that l_i corresponds to m_i and not to m_j for $j \neq i$).

It is now easily seen that the two Johansen test statistics given by (20) and (6) are the same, since substituting (13) into (6) yields

$$T \sum_{i=1}^{G-r} \log(l_i) = T \sum_{i=1}^{G-r} \log\left(\frac{1}{1-m_i}\right) = -T \sum_{i=1}^{G-r} \log(1-m_i), \qquad \text{Q.E.D.}$$

We conclude that the Johansen test with null-hypothesis "there exist at most $G-1$ cointegrating relations" can be interpreted as an Anderson-Rubin test with null-hypothesis "the variables in the vector y_{t-1} can be excluded from a structural equation with endogenous variables Δy_t and predetermined variables $\Delta y_{t-1}, \ldots, \Delta y_{t-k+1}$ and D_t". If it would be allowed to apply the Anderson-Rubin test to this model it would yield exactly the same value of the test statistic as the Johansen test statistic with the same p-value. However in the (IN)SEM no nonstationary variables are allowed to occur, so that we would have to compare the value of this 'Anderson-Rubin' test statistic with a 'non-standard' critical value, given by Johansen (1995).

Since we now know that the test statistics (20) and (6) are the same functions of the data, we might expect that we could also use the procedure that is commonly used for obtaining the Johansen test statistic (in appendix A) in order to compute the Anderson-Rubin test statistic, if we again let appropriate variables play the roles of Z_{0t}, Z_{1t} and Z_{2t}. However, in the case of the Anderson-Rubin test the proof of the equivalence of the procedures is a little different, as in this case the parameter matrix with reduced rank is not always square. If $k_0 > g_1 + 1$ the $(g_1+1) \times k_0$ matrix A^* and the $k_0 \times (g_1+1)$ matrix B^* in (11) are not square, so that it follows from Lemma 1 that the eigenvalues of the $k_0 \times k_0$ matrix

$$B^*A^* = S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$$

in (12) consist of the eigenvalues of the $(g_1+1) \times (g_1+1)$ matrix

$$A^*B^* = S_{00}^{-1/2} S_{01} S_{11}^{-1} S_{10} S_{00}^{-1/2}$$

plus $k_0 - (g_1+1)$ times the eigenvalue zero. In order to have that $l_i = 1/(1-m_i)$ (for $i = 1, \ldots, g_1+1$) we need to define the m_i's as the *nonzero* eigenvalues of (12). The alternative procedure for obtaining the Anderson-Rubin test statistic is therefore given by:

1. Regress $(y_{1t}, Y_{1t'})'$ and X_{0t} on X_{1t}, and obtain the residual vectors r_{0t} and r_{1t} ($t = 1, \ldots, T$), respectively.
2. Compute the matrices S_{ij}, which are defined as:

$$S_{ij} \equiv \frac{1}{T} \sum_{t=1}^{T} r_{it} r_{jt}' \qquad (i,j = 0,1).$$

3. Compute the eigenvalues of the $k_0 \times k_0$ matrix

$$S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2},$$

and define the m_i's as the g_1+1 *nonzero* eigenvalues with $0 < m_1 < m_2 < \ldots < m_{g_1+1} < 1$. Then the likelihood ratio test statistic is

$$LR = -T \log(1-m_1). \tag{14}$$

A small advantage of this alternative procedure is that (for moderate and large samples) it is faster than the conventional Anderson-Rubin procedure, because we now only need to compute one inverse matrix, $(X_1'X_1)^{-1}$, in the regressions in step 1 instead of both $(X_1'X_1)^{-1}$ and $(X'X)^{-1}$. For example, in a sample of size 80 with $(y_{1t}, Y_{1t}')'$, X_{0t} and X_{1t} all of size 4×1 the alternative procedure is approximately 1.6 times as fast as the conventional procedure (using Gauss Light, version 3.2.31).

4 The Likelihood Ratio Test for Reduced Rank in the General Case

We have now seen that the Anderson-Rubin test statistic and the Johansen test statistic can be obtained by very similar procedures. One might expect that in the general case the likelihood ratio test for reduced rank can always be performed in such a way. The likelihood ratio statistic in a test of $H_0 : \text{rank}(A_1) \leq r$ against $H_1 : \text{rank}(A_1) \geq r + 1$ in the general reduced rank regression model (1) can indeed be obtained in the following way:

1. Regress Z_{0t} and Z_{1t} on Z_{2t}, and obtain the residual vectors r_{0t} and r_{1t} $(t = 1, \ldots, T)$, respectively.
2. Compute the matrices S_{ij}, which are defined as:

$$S_{ij} \equiv \frac{1}{T} \sum_{t=1}^{T} r_{it} r_{jt}' \qquad (i, j = 0, 1). \tag{15}$$

3. Compute the n_1 eigenvalues of the matrix

$$S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2},$$

and define the m_i's as the n nonzero eigenvalues with $0 < m_1 < m_2 < \ldots < m_n < 1$, where we define n as $n \equiv \min\{n_0, n_1\}$, so that A_1 has full rank if $\text{rank}(A_1) = n$.
Then the likelihood ratio test statistic is

$$LR = -T \sum_{i=1}^{n-r} \log(1 - m_i). \tag{16}$$

If the explanatory variables in the model are stationary, the (asymptotic) distribution of LR under H_0 is χ^2 with $(n_0 - r)(n_1 - r)$ degrees of freedom.
If the explanatory variables are nonstationary, the asymptotic distribution of LR under H_0 is non-standard.

The result is formalized in the following theorem:

Theorem 2. *The test statistic* (16) *is the likelihood ratio statistic for the test of* $H_0 : \text{rank}(A_1) \leq r$ *against* $H_1 : \text{rank}(A_1) \geq r + 1$ *in the general reduced rank regression model* (1).

The proof of the theorem is given in appendix B. Not surprisingly, this proof is similar to the derivation of the Johansen test in Johansen (1995).

5 Singular Value Interpretation

We shall now show that both the Anderson-Rubin test and the Johansen test can be interpreted as a singular value decomposition. Let us again consider the general reduced rank regression model (1):

$$Z_{0t} = A_1' Z_{1t} + A_2' Z_{2t} + u_t \quad (t = 1, \ldots, T).$$

The (unrestricted) OLS estimator of the matrix A_1, of which the role is played by (π_{10}, Π_{10}) in the INSEM and by Π in the VAR model, is given by

$$\hat{A}_1 = S_{11}^{-1} S_{10},$$

because of the Frisch-Waugh theorem for partial regression (and the definition of the S_{ij}'s in (15)). Let the matrix $\tilde{A}_1$ be defined as the following 'scaled' version of the matrix $\hat{A}_1$

$$\tilde{A}_1 = S_{11}^{1/2} \hat{A}_1 \hat{\Omega}^{-1/2} = S_{11}^{-1/2} S_{10} \hat{\Omega}^{-1/2}, \tag{17}$$

which we can interpret as the matrix $\hat{A}_1$ scaled by the square root of its 'covariance matrix' $S_{11}^{-1} \otimes \hat{\Omega}$, see also Johansen (1995). The likelihood ratio statistic can now be formulated in terms of the singular values of the matrix $\tilde{A}_1$, which is formalized in the following theorem:

Theorem 3. *The likelihood ratio statistic*

$$LR = -T \sum_{i=1}^{n-r} \log(1 - m_i).$$

in the test of $H_0 : \text{rank}(A_1) \leq r$ *against* $H_1 : \text{rank}(A_1) \geq r + 1$ *in the model* (1) *can be written as:*

$$LR = T \sum_{i=1}^{n-r} \log(1 + s_i^2),$$

where the s_i*'s (with* $s_1 \leq \ldots \leq s_n$*) are the singular values of the matrix* $\tilde{A}_1$ *in* (17) *where* $n \equiv min\{n_0, n_1\}$.

Proof: We first notice that the squared singular values s_i^2 are defined as the n (nonzero) eigenvalues of the matrix

$$\tilde{A}_1' \tilde{A}_1 = \hat{\Omega}^{-1/2} S_{01} S_{11}^{-1} S_{10} \hat{\Omega}^{-1/2},$$

which are given by the equation

$$|s^2 I - \hat{\Omega}^{-1/2} S_{01} S_{11}^{-1} S_{10} \hat{\Omega}^{-1/2}| = 0 \quad \Leftrightarrow \quad |s^2 \hat{\Omega} - S_{01} S_{11}^{-1} S_{10}| = 0.$$

Substituting (9) yields

$$|s^2 (S_{00} - S_{01} S_{11}^{-1} S_{10}) - S_{01} S_{11}^{-1} S_{10}| = 0,$$

which is equivalent with

$$\begin{aligned} \left| s^2 S_{00} - (1 + s^2) S_{01} S_{11}^{-1} S_{10} \right| = 0 \quad &\Leftrightarrow \quad \left| \frac{s^2}{1 + s^2} S_{00} - S_{01} S_{11}^{-1} S_{10} \right| = 0 \\ &\Leftrightarrow \quad \left| \frac{s^2}{1 + s^2} I - S_{00}^{-1/2} S_{01} S_{11}^{-1} S_{10} S_{00}^{-1/2} \right| = 0, \end{aligned}$$

so that for the singular values s_i it holds true that $s_i^2/(1 + s_i^2)$ is one of the nonzero eigenvalues of

$$S_{00}^{-1/2} S_{01} S_{11}^{-1} S_{10} S_{00}^{-1/2},$$

which are by Lemma 1 (with $A = S_{00}^{-1/2} S_{01} S_{11}^{-1/2}$ and $B = S_{11}^{-1/2} S_{10} S_{00}^{-1/2}$) equal to the nonzero eigenvalues m_i $(i = 1, \ldots, n)$ of

$$S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}.$$

It follows that

$$\frac{1}{1 - m_i} = \frac{1}{1 - \left(\frac{s_i^2}{1 + s_i^2} \right)} = \frac{1 + s_i^2}{1 + s_i^2 - s_i^2} = 1 + s_i^2 \qquad (i = 1, \ldots, n).$$

This implies that

$$-T \log(1 - m_i) = T \log(1 + s_i^2) \qquad (i = 1, \ldots, n),$$

so that

$$-T \sum_{i=1}^{n-r} \log(1 - m_i) = T \sum_{i=1}^{n-r} \log(1 + s_i^2),$$

Q.E.D.

Since the reduced form of the (IN)SEM and the VECM are both instances of the general reduced rank regression model in (1), we have that both the Anderson-Rubin test and the Johansen test have the interpretation of a singular value decomposition. The Anderson-Rubin test statistic can be interpreted as a monotonically increasing function of the the smallest singular value of the 't-value' of the (unrestricted) least-squares estimator of (π_{10}, Π_{10}). The Johansen test statistic can be interpreted as a monotonically increasing function of the $G - r$ smallest singular values of the 't-value' of the (unrestricted) least-squares estimator of Π.

6 Conclusion

The most important similarity of the Anderson-Rubin test and the Johansen test is that these tests are applied to models with a very similar mathematical structure: the reduced form of the (IN)SEM under an overidentification restriction and the VECM are both instances of a general reduced rank regression model

$$Z_{0t} = A_1' Z_{1t} + A_2' Z_{2t} + u_t \qquad (t = 1, \ldots, T),$$

in which both tests are likelihood ratio tests of H_0: $\text{rank}(A_1) \leq r$ against H_1: $\text{rank}(A_1) \geq r + 1$ for some integer r. As might be expected, the procedure that is commonly used for obtaining the Johansen test statistic can be used in order to obtain the Anderson-Rubin test statistic, and vice versa. Another similarity is that both statistics are monotonically increasing functions of the singular values of a scaled version, the 't-value', of the (unrestricted) least-squares estimator of the matrix A_1.

A The Conventional Procedures for Obtaining the Anderson-Rubin and Johansen Test Statistics

The Anderson-Rubin test is the likelihood ratio test of $H_0 : \text{rank}(\pi_{10}, \Pi_{10}) \leq g_1$ against $H_1 : \text{rank}(\pi_{10}, \Pi_{10}) = g_1 + 1$ in the reduced form of the (IN)SEM (3), and has the interpretation of a likelihood ratio test of the validity of the restrictions on the first structural equation of the (IN)SEM. Anderson and Rubin (1949) showed that the likelihood ratio statistic can be obtained by the following procedure:

1. Regress the endogenous variables $(y_{1t} \;\; Y_{1t}')'$ on X_{1t}, the predetermined variables occurring in the first structural equation, and obtain the residual vectors e_t^* $(t = 1, \ldots, T)$. The sample covariance matrix of these residuals is given by

$$\hat{\Omega}^* \equiv \frac{1}{T} \sum_{t=1}^{T} e_t^* e_t^{*'}.$$

2. Regress the endogenous variables $(y_{1t} \quad Y_{1t}')'$ on $X_t = (X_{1t}' \quad X_{0t}')'$, all predetermined variables occurring in the model, and obtain the residual vectors e_t. The sample covariance matrix of these residuals is

$$\hat{\Omega} \equiv \frac{1}{T} \sum_{t=1}^{T} e_t e_t'.$$

3. Determine the eigenvalues l_i $(i = 1, \ldots, g_1 + 1)$ of the matrix

$$\hat{\Omega}^{-1/2} \hat{\Omega}^* \hat{\Omega}^{-1/2}, \tag{18}$$

where $1 \leq l_1 \leq l_2 \leq \ldots \leq l_{g_1+1}$. Then the likelihood ratio test statistic is given by

$$LR = T \log(l_1),$$

which is asymptotically $\chi^2_{k_0 - g_1}$ distributed under H_0.

Notice that the Anderson-Rubin test can only be applied if $k_0 > g_1$, which is called the case of 'overidentification' since more than enough exogenous variables are omitted from the first structural equation to make identification of γ_1 and β_1 possible (see Greene (1997)). The Anderson-Rubin test can therefore only be used to examine whether the 'overidentifying' restrictions on the structural form are supported by the data: it is a test of 'overidentification'.

The Johansen test is the likelihood ratio test of $H_0 : \text{rank}(\Pi) \leq r$ against $H_1 : \text{rank}(\Pi) \geq r+1$ in the VAR model in error correction form (4), and has the interpretation of a likelihood ratio test of the validity of the cointegration restrictions that the VECM imposes on the VAR model. Johansen (1991) showed that the likelihood ratio statistic can be obtained by the following procedure:

1. Regress Δy_t and y_{t-1} on Δy_{t-1}, ..., Δy_{t-k+1}, and D_t, and obtain the residual vectors r_{0t} and r_{1t} $(t = 1, \ldots, T)$, respectively.

2. Compute the matrices S_{ij}, which are defined as:

$$S_{ij} \equiv \frac{1}{T} \sum_{t=1}^{T} r_{it} r_{jt}' \qquad (i, j = 0, 1).$$

3. Compute the eigenvalues m_i $(i = 1, \ldots, G)$ of the matrix

$$S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}, \tag{19}$$

where $0 < m_1 < m_2 < \ldots < m_G < 1$. Then the likelihood ratio test statistic is given by

$$LR = -T \sum_{i=1}^{G-r} \log(1 - m_i). \tag{20}$$

The (asymptotic) distribution of LR under H_0 is non-standard, since under H_0 (i.e. in the VECM) some of the explanatory variables in the model are nonstationary. Some critical values can be found in Johansen (1995).

B Derivation of the Likelihood Ratio Test for Reduced Rank in the General Case

The proof of Theorem 2 on page 7 is as follows. Under the restriction $\text{rank}(A_1) \leq r$ the $n_0 \times n_1$ matrix A_1' can be factorized as the product

$$A_1' = \alpha\beta',$$

where α is an $n_0 \times r$ matrix and β is an $n_1 \times r$ matrix. If we substitute this into (1), the general reduced rank regression model becomes

$$Z_{0t} = \alpha\beta' Z_{1t} + A_2' Z_{2t} + u_t, \quad t = 1, \ldots, T.$$

Since the u_t's are I.I.D. $N(0, \Omega)$ distributed, the log-likelihood function is given by

$$\begin{aligned} \log L(\alpha, \beta, A_2, \Omega) &= -\frac{Tn_0}{2}\log(2\pi) - \frac{T}{2}\log(\det(\Omega)) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(Z_{0t} - \alpha\beta' Z_{1t} - A_2' Z_{2t})'\Omega^{-1} \times \\ &\qquad (Z_{0t} - \alpha\beta' Z_{1t} - A_2' Z_{2t}). \end{aligned}$$

The $n_0 n_1$ first order conditions for estimating A_2 are given by

$$\sum_{t=1}^{T}(Z_{0t} - \alpha\beta' Z_{1t} - \hat{A}_2' Z_{2t}) Z_{2t}' = 0.$$

In terms of the product moment matrices M_{ij} $(i, j = 0, 1, 2)$, defined as

$$M_{ij} \equiv \frac{1}{T}\sum_{t=1}^{T} Z_{it} Z_{jt}',$$

the first order conditions are given by

$$M_{02} = \alpha\beta' M_{12} + \hat{A}_2' M_{22},$$

so that for given values of α and β the maximum likelihood estimator $\hat{A}_2$ is given by

$$\hat{A}_2(\alpha, \beta) = M_{22}^{-1} M_{20} - M_{22}^{-1} M_{21}\beta\alpha'. \tag{21}$$

Defining the vectors r_{0t} and r_{1t} $(t = 1, \ldots, T)$ as the residuals in the regressions of Z_{0t} and Z_{1t} on Z_{2t}, we have that

$$\begin{aligned} r_{0t}' &= Z_{0t}' - Z_{2t}' M_{22}^{-1} M_{21}, \\ r_{1t}' &= Z_{1t}' - Z_{2t}' M_{22}^{-1} M_{21}. \end{aligned}$$

Since

$$Z_{0t} - \alpha\beta' Z_{1t} - \hat{A}_2' Z_{2t} = r_{0t} - \alpha\beta' r_{1t},$$

the concentrated log-likelihood function in terms of α, β and Ω is

$$\begin{aligned} L(\alpha, \beta, \Omega) &= -\frac{Tn_0}{2}\log(2\pi) - \frac{T}{2}\log(\det(\Omega)) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(r_{0t} - \alpha\beta' r_{1t})'\Omega^{-1}(r_{0t} - \alpha\beta' r_{1t}). \end{aligned} \tag{22}$$

A regression equation in the residuals

$$r_{0t} = \alpha\beta' r_{1t} + \tilde{\varepsilon}_t \qquad (t = 1, \ldots, T) \tag{23}$$

would give the same log-likelihood as (22). We conclude that the parameters A_2 can be eliminated by regression.

For fixed β it is easy to estimate α and Ω by regressing r_{0t} on $\beta' r_{1t}$ and obtaining (using the definition of S_{ij} in (15))

$$\begin{aligned} \hat{\alpha}(\beta) &= S_{01}\beta(\beta' S_{11}\beta)^{-1}, & (24)\\ \hat{\Omega}(\beta) &= S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} \\ &= S_{00} - \hat{\alpha}(\beta)(\beta' S_{11}\beta)\hat{\alpha}(\beta)'. & (25) \end{aligned}$$

Defining the $T \times n_0$ matrix R_0 and the $T \times n_1$ matrix R_1 as the matrices with the rows r_{0t}' and r_{1t}' $(t = 1, \ldots, T)$, respectively, we write the last term of (22) as

$$-\frac{1}{2}\sum_{t=1}^{T}(r_{0t} - \alpha\beta' r_{1t})'\Omega^{-1}(r_{0t} - \alpha\beta' r_{1t}) =$$

$$-\frac{1}{2}\mathrm{tr}((R_0' - \alpha\beta' R_1')'\Omega^{-1}(R_0' - \alpha\beta' R_1')). \tag{26}$$

Using the property of the trace that $\mathrm{tr}(QR) = \mathrm{tr}(RQ)$ we write (26) as

$$-\frac{1}{2}\mathrm{tr}(\Omega^{-1}(R_0' - \alpha\beta' R_1')(R_0' - \alpha\beta' R_1')') =$$

$$-\frac{T}{2}\mathrm{tr}(\Omega^{-1}(S_{00} - \alpha\beta' S_{11}\beta\alpha' - \alpha\beta' S_{10} - S_{01}\beta\alpha')). \tag{27}$$

If we substitute (24) and (25) for α and Ω in (27), we have

$$-\frac{T}{2}\mathrm{tr}((S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10})^{-1}(S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10})) =$$

$$= -\frac{T}{2}\mathrm{tr}(I_{n_0}) = -\frac{Tn_0}{2}. \tag{28}$$

This implies that the concentrated log-likelihood function in terms of β is given by

$$\begin{aligned} L(\beta) &= -\frac{Tn_0}{2}(\log(2\pi) + 1) - \frac{T}{2}\log(\det(\hat{\Omega}(\beta))) \\ &= -\frac{Tn_0}{2}(\log(2\pi) + 1) \\ &\quad -\frac{T}{2}\log(\det(S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10})). & (29) \end{aligned}$$

In order to rewrite this expression we use the identity

$$\begin{aligned} \det\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} &= \det(\Sigma_{11})\det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \\ &= \det(\Sigma_{22})\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \end{aligned}$$

Applying the identity to the matrix

$$\begin{pmatrix} S_{00} & S_{01}\beta \\ \beta' S_{10} & \beta' S_{11}\beta \end{pmatrix}$$

we have

$$\begin{aligned}\det\begin{pmatrix} S_{00} & S_{01}\beta \\ \beta' S_{10} & \beta' S_{11}\beta \end{pmatrix} &= \det(S_{00})\det(\beta'(S_{11}-S_{10}S_{00}^{-1}S_{01})\beta) \\ &= \det(\beta' S_{11}\beta)\times \\ &\quad \det(S_{00}-S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}),\end{aligned}$$

so that

$$\det(S_{00}-S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}) = \det(S_{00})\frac{\det(\beta'(S_{11}-S_{10}S_{00}^{-1}S_{01})\beta)}{\det(\beta' S_{11}\beta)}. \tag{30}$$

The maximization of the likelihood function is therefore equivalent with the minimization of the last factor of (30). We now give a lemma that we shall use for this minimization.

Lemma 2. *Let M be a positive semi-definite symmetric $G\times G$ matrix and N a positive definite symmetric $G\times G$ matrix. Then the function*

$$f(x) = \frac{\det(x'Mx)}{\det(x'Nx)}$$

is minimized among all $G\times r$ matrices x by $\hat{x}=(v_1,\ldots,v_r)$, and the minimal value is $\prod_{i=1}^{r}\lambda_i$, where λ_i and v_i are solutions to the eigenvalue problem

$$\det(\lambda N - M) = 0 \quad\Leftrightarrow\quad \lambda_i N v_i = M v_i$$

where $0<\lambda_1\leq\ldots\leq\lambda_G$. We can also choose $\hat{x}$ times any nonsingular $r\times r$ matrix as the minimizing value.

For the proof of this lemma we refer to Johansen (1995). It follows from Lemma 2 that the last factor of (30) is minimized among all $n_1\times r$ matrices β by solving the eigenvalue problem

$$\det(kS_{11}-(S_{11}-S_{10}S_{00}^{-1}S_{01})) = 0,$$

i.e. by determining the r smallest eigenvalues k_i and the corresponding eigenvectors v_i. For $h=1-k$, this is equivalent with solving the eigenvalue problem

$$\det(hS_{11}-S_{10}S_{00}^{-1}S_{01}) = 0,$$

i.e. we have to compute the r *largest* eigenvalues $h_i=1-k_i$ ($h_{n_1-r+1}\leq\ldots\leq h_{n_1}$), and the corresponding eigenvectors v_i, that satisfy the equation

$$S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}v_i = h_i v_i.$$

The estimator of β is then given by

$$\hat{\beta} = (v_{n_1-r+1},\ldots,v_{n_1}), \tag{31}$$

the $n_1\times r$ matrix of which the columns are the eigenvectors of $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ corresponding to the r largest eigenvalues.

With the choice of $\hat{\beta}$ in (31) we find from Lemma 2 that

$$\begin{aligned}L(\hat{\beta}) &= -\frac{Tn_0}{2}(\log(2\pi)+1) \\ &\quad -\frac{T}{2}\log\left(\det(S_{00})\frac{\det(\hat{\beta}'(S_{11}-S_{10}S_{00}^{-1}S_{01})\hat{\beta})}{\det(\hat{\beta}' S_{11}\hat{\beta})}\right)\end{aligned}$$

$$
\begin{aligned}
&= -\frac{Tn_0}{2}(\log(2\pi)+1) - \frac{T}{2}\log\left(\det(S_{00})\prod_{i=n_1-r+1}^{n_1}(1-h_i)\right) \\
&= -\frac{Tn_0}{2}(\log(2\pi)+1) - \frac{T}{2}\log(\det(S_{00})) \\
&\quad -\frac{T}{2}\sum_{i=n_1-r+1}^{n_1}\log(1-h_i). \qquad (32)
\end{aligned}
$$

Let us denote the general reduced rank regression model (1) under the restriction $\text{rank}(A_1) \leq r$ by $H(r)$. For each r the maximized log-likelihood value is given by (32). Subtracting the maximized log-likelihood function for $H(r)$ from the corresponding expression for the full rank case $H(n)$ yields the logarithm of the likelihood ratio:

$$
\begin{aligned}
L_{H(r)} - L_{H(n)} &= -\frac{T}{2}\left(\sum_{i=n_1-r+1}^{n_1}\log(1-h_i) - \sum_{i=n_1-n+1}^{n_1}\log(1-h_i)\right) \\
&= \frac{T}{2}\sum_{i=n_1-n+1}^{n_1-r}\log(1-h_i),
\end{aligned}
$$

so that the likelihood ratio statistic for testing the hypothesis or submodel $H(r)$ in $H(n)$ is given by

$$
LR(r) = -2(L_{H(r)} - L_{H(n)}) = -T\sum_{i=n_1-n+1}^{n_1-n+n-r}\log(1-h_i). \qquad (33)
$$

Since the $n_1 \times n_1$ matrix $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ has rank n (with probability 1), it has the eigenvalue 0 with multiplicity $n_1 - n$ if $n_1 > n$. Defining the m_i's (for $i = 1, \ldots, n$) as the *nonzero* eigenvalues h_i $(i = n_1 - n + 1, \ldots, n_1)$, we have that

$$
LR(r) = -T\sum_{i=1}^{n-r}\log(1-m_i).
$$

So we have that the likelihood ratio test statistic is a function of the r smallest nonzero eigenvalues of the matrix $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$, which are by Lemma 1 (with $A = S_{11}^{-1/2}$ and $B = S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}$) equal to the eigenvalues of the matrix $S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}$. We conclude that the likelihood ratio statistic is given by (16) on page 7, Q.E.D.

References

[1] Anderson, T.W. and H. Rubin (1949), Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics*, **20**.

[2] Greene, W.H. (1997), *Econometric Analysis*, third ed., Prentice Hall, Upper Saddle River, New Jersey.

[3] Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica* **59**.

[4] Johansen, S. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, first ed., Oxford University Press, Oxford.

[5] Koopmans, T.C. and W.C. Hood (1953), *Estimation of Simultaneous Linear Economic Relationships*, Cowles Commission Monograph. No. 14. John Wiley & Sons, New York.